Report Number— NPS63-80-002

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

EXPERIMENTS CONCERNING CATEGORICAL FORECASTS OF OPEN-OCEAN VISIBILITY USING MODEL OUTPUT STATISTICS

by

Philip George Yavorsky

June 1980

Thesis Advisor: R. J. Renard

Approved for public release; distribution unlimited.

Prepared for:

Naval Postgraduate School
Monterey, California

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral J. J. Ekelund
Superintendent

Jack R. Borsting
Provost

This thesis prepared in conjunction with research supported by the Naval Air Systems Command (AIR 370).

Released as a
Technical Report by

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPS63-80-002	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Experiments Concerning Categorical Forecasts of Open-Ocean Visibility Using Model Output Statistics		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; June 1980
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Philip George Yavorsky and Robert J. Renard		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE June 1980
		13. NUMBER OF PAGES 87
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

North Pacific Ocean visibility
Visibility
Model Output Statistics
Visibility Forecasting
Marine Visibility
Marine Fog

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This study is an extension of previous statistically oriented research at the Naval Postgraduate School. The method of Model Output Statistics is used to predict open-ocean visibility employing stepwise-selection, multiple linear regression. The visibility predictand is specified categorically with comparisons made to a previous probabilistic approach. Predictors include direct and derived model output parameters

(20. ABSTRACT Continued)

provided by the U.S. Navy's Fleet Numerical Oceanography Center (FNOC), Monterey, California. About 18,000 North Pacific Ocean (30°-60°N) synoptic ship reports at 0000 GMT from June 1976 and 1977, July 1979, and August 1979 were used as both dependent and independent data sets. Visibility equations for both analysis-time and 24- and 48-hr prognostic times are developed, and are verified using percent correct, Heidke skill score and bias. Levels of skill are less than desirable for operational use. Important predictor parameters are found to be sensible and evaporative heat fluxes, meridional wind component, sea-level pressure, air/sea temperature difference, relative humidity, an FNOC fog probability parameter and a visibility parameter derived from a marine aerosol model. Other experiments concerning weighted least squares, predictand transformations and R^2 deflation are briefly described.

Experiments Concerning Categorical Forecasts of Open-Ocean Visibility Using Model Output Statistics

by

Philip George Yavorsky
Captain, United States Air Force
B.S., United States Air Force Academy, 1973
B.S., University of Utah, 1974

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL

June 1980

ABSTRACT

This study is an extension of previous statistically oriented research at the Naval Postgraduate School. The method of Model Output Statistics is used to predict open-ocean visibility employing stepwise-selection, multiple linear regression. The visibility predictand is specified categorically with comparisons made to a previous probabilistic approach. Predictors include direct and derived model output parameters provided by the U.S. Navy's Fleet Numerical Oceanography Center (FNOC), Monterey, California. About 18,000 North Pacific Ocean (30°-60°N) synoptic ship reports at 0000 GMT from June 1976 and 1977, July 1979, and August 1979 were used as both dependent and independent data sets. Visibility equations for both analysis-time and 24- and 48-hr prognostic times are developed, and are verified using percent correct, Heidke skill score, and bias. Levels of skill are less than desirable for operational use. Important predictor parameters are found to be sensible and evaporative heat fluxes, meridional wind component, sea-level pressure, air/sea temperature difference, relative humidity, an FNOC fog probability parameter and a visibility parameter derived from a marine aerosol model. Other experiments concerning weighted least squares predictand transformations and R^2 deflation are briefly described.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

There are many people to whom I desire to express my gratitude for help on this thesis. Above all, I am thankful for the guidance of my thesis advisor, Dr. Robert J. Renard, whose constant interest and pleasant personality made this research both challenging and enjoyable.

I also extend my gratitude to Dr. Willem van der Bijl for his aid in understanding statistical concepts, ensuring proper interpretation of statistical results, and his occasional humor. Thanks also go to Dr. Donald P. Gaver for his statistical insight which provided many new avenues to investigate. And worthy of mention is Dr. Andrzej Goroch of the Naval Environmental Prediction Research Facility, whom I thank for his friendly cooperation in providing the information and explanation needed to compute the beta visiblity parameter.

Special thanks go to Mr. William J. Thompson for his friendship, his aid in solving programming problems, and for spending many extra hours preparing the July and August 1979 data sets. I also appreciate the efforts of Lt. William T. Aldinger who helped to introduce me to this research, and who prepared the June 1976 and 1977 data sets.

And, of course, I must thank the staff of the W.R. Church Computer Center, many of whom helped me solve computer problems, especially Mr. Dennis Mar; and Ms. Kristina Butler who allowed

the project special computer priority so that results could be ready for a conference presentation.

Final thanks go to Mr. Michael McDermet and his staff, NPS, for their willingness to assist in solving a number of technical problems encountered in the thesis work.

I. INTRODUCTION AND BACKGROUND

Visibility is an important meteorological variable that can have a significant impact on the safety of maritime operations. Naval activities such as amphibious assault, underway replenishment and air operations can be greatly restricted under conditions of low visibility. Civilian operations can suffer also. In most cases poor visibility at sea is due to the occurrence of fog. The economic, military and human losses associated with United States Naval Operations attributable to fog are well documented by Wheeler and Leipper (1974). Thus accurate forecasts of fog, or more generally, marine visibility, would be of great benefit to the military and civilian communities.

Earlier research into this problem at the Naval Postgraduate School (NPS), Monterey, California, using statistical methods, was conducted by Van Orman and Renard (1977), Quinn (1978), and Ouzts and Renard (1979), who all applied regression techniques to forecast the occurrence of fog with some degree of skill. Research into forecasting visibility directly, but using a very limited set of parameters and data, was conducted by Schramm (1966). Further work by Nelson (1972) used a larger data set and investigated new parameters. More recently the work by Aldinger (1979) continued research into determining those parameters which are statistically correlated with marine visibility. In addition, using a probabilistic approach,

Aldinger derived analysis-time linear regression equations which show a reasonable degree of probabilistic skill. He also expanded the evaluation of these equations to categorical estimates using Threat Score, Heidke Skill Score and percent correct. In addition, he adapted a scoring awards matrix to the verification which enhances the skill by giving partial credit to forecasts that are close to the observed category.

This study continues the statistical regression work on visibility analysis/forecasting, but uses a categorical approach rather than a probabilistic one. New predictor parameters are investigated and prognostic, as well as analysis-time, equations are derived. In addition, more attention is given to interpreting the statistical methods used.

II. OBJECTIVES

The primary objective of this study was to expand on previous NPS visibility research using numerical-model output parameters from the Fleet Numerical Oceanography Center (FNOC),[1] Monterey, California to diagnose and predict marine visibility over the open ocean by statistical means. The method of model output statistics (MOS) (see Glahn and Lowry, 1972) was used to predict visibility categories directly as opposed to using a probabilistic approach.

Within the primary objective, more specific goals to be achieved were to:

(1) Develop statistical diagnostic (analysis-time, or Tau Ø hr) and prognostic (forecast-time, or Tau 24 hr, 48 hr) visibility equations using stepwise multiple linear regression;

(2) test several types of categorical schemes;

(3) test various forms of the visibility predictand in the regression program;

(4) test predictor parameters not previously used in NPS visibility research;

(5) compare the categorical approach to the probabilistic approach as used by Aldinger (1979);

(6) test methods of regression other than the least-squares linear type.

---

[1]Formerly called the "Fleet Numerical Weather Central".

III. DATA

A. AREA

The area of study was limited to a region of the North Pacific Ocean located approximately between 30° and 60°N and from 145°E to 130°W. The actual area was restricted in size from the limits mentioned in order to reduce the number of land-influenced grid points used in computing derivatives applicable at marine grid locations. Also, this was done to eliminate, as much as possible, any orographic influences on visibility. The study area is shown in Figure 1 on a polar stereographic projection, the grid points of which correspond to those of the standard FNOC 63 x 63 grid (with a mesh size of 381 km at 60°N). The entire FNOC grid is shown in Figure 2 with an outlined area from which FNOC's model output parameters were extracted. This study area is the same as that used for recent statistical studies of marine fog and visibility at NPS.

B. SELECTION OF TIME PERIOD

Data from the months of June, July and August only were used in this study. The frequency of fog - (and thus visibility) related maritime casualties reaches a peak during the Northern Hemisphere summer months (Figure 3). Therefore, this period is one of primary operational significance.

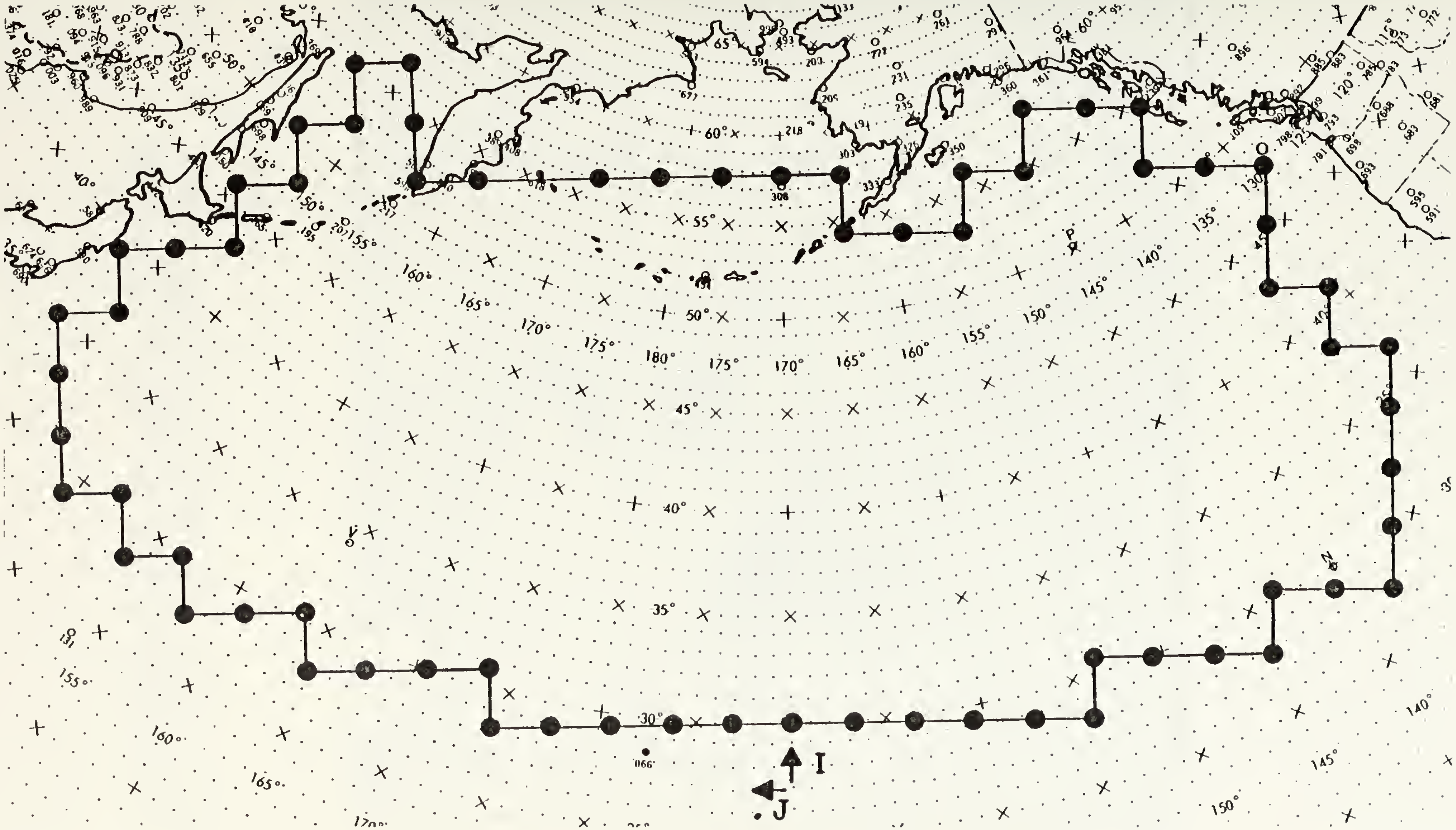


Figure 1. Study area on polar stereographic projection.

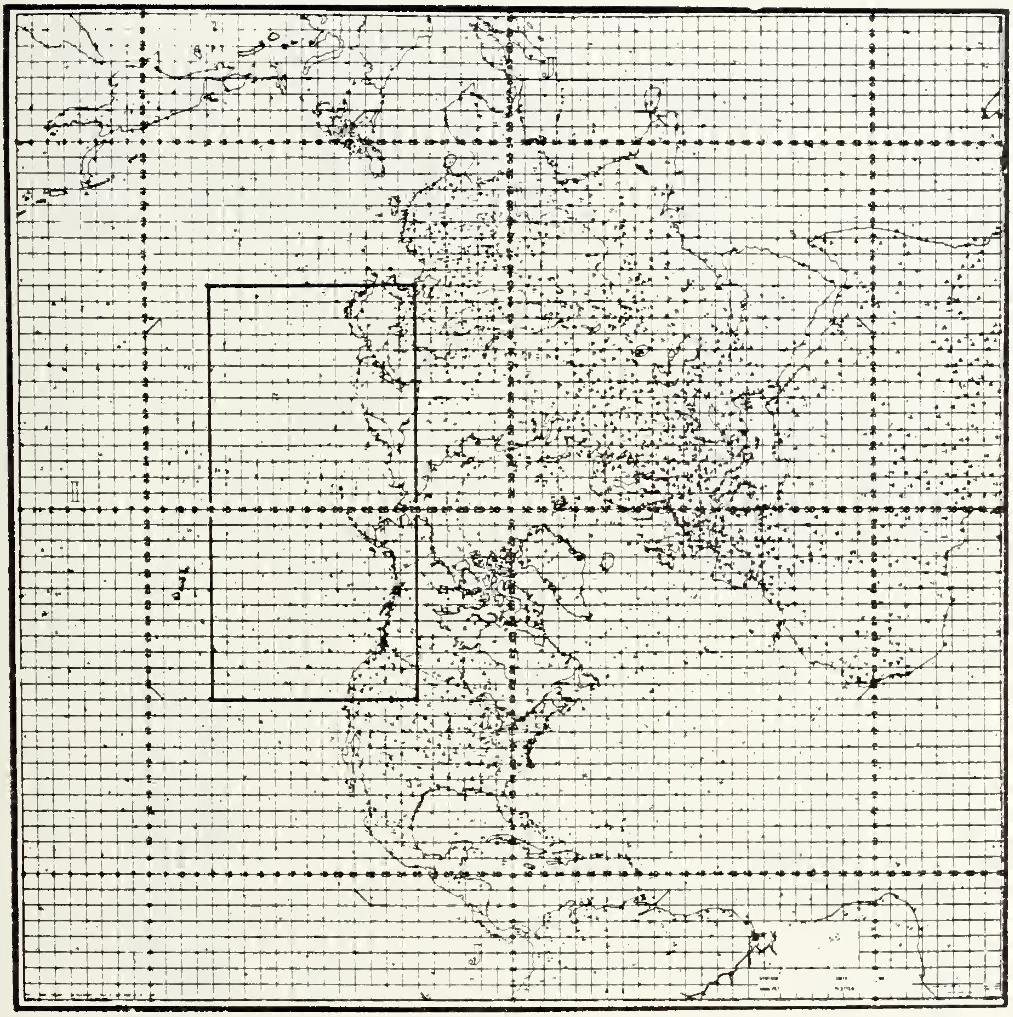


Figure 2. Fleet Numerical Oceanography Center's 63 x 63 grid, with outline of North Pacific Ocean rectangular grid area used in study.

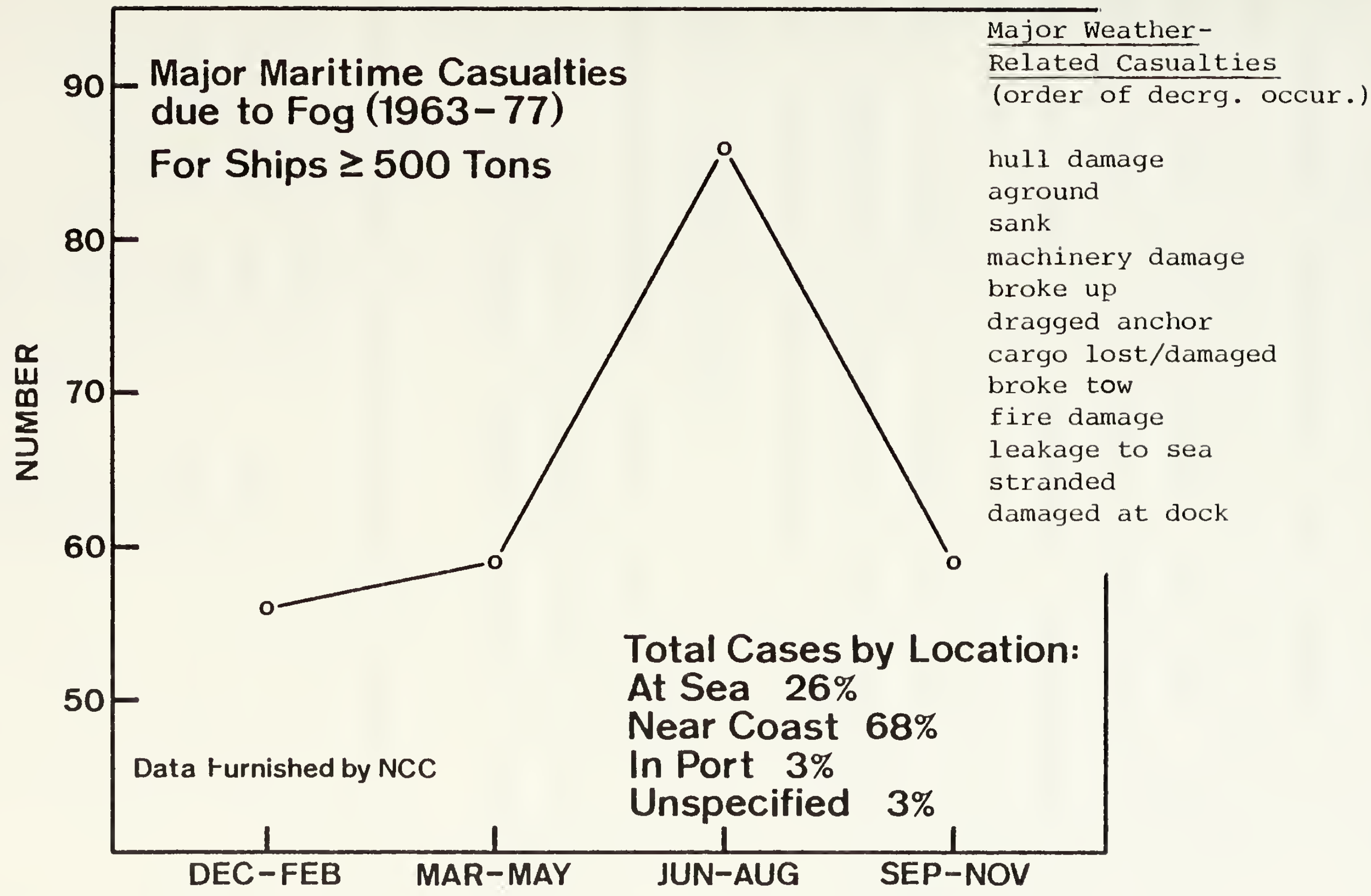


Figure 3. Major Maritime Casualties due to fog (1963-77) for ships $\geq$ 500 tons.

Only 0000 GMT synoptic ship report data were used as this ensured that daylight was present throughout the study area, thus allowing more accurate visibility observations than if nighttime observations were included.

Model output parameter data from FNOC were taken from 0000 GMT for use in analysis-time equations. However, in prognostic equations 1200 GMT parameters also were used.

Diagnostic (Tau 0 hr) equations were developed from combined June 1976 and June 1977 data using analysis-time data only. In addition, equations for Tau 0, 24 and 48 hrs were developed from July 1979 data using both analysis-time and prognostic-time parameters.

C. SYNOPTIC WEATHER REPORTS

The synoptic weather reports used in this study were provided by the Naval Oceanography Command Detachment[2] co-located with the National Climatic Center at Asheville, North Carolina.

The total number of observations available in the area of Figure 1 is as follows:

June 1976	(Tau 0)	4277
June 1977	(Tau 0)	5044
July 1979	(Tau 0)	4079
	(Tau 24)	4095
	(Tau 48)	4102

---

[2]Formerly called the "Naval Weather Service Detachment".

August 1979	(Tau 0)	4727
	(Tau 24)	4520
	(Tau 48)	4421

The actual number of cases varied slightly from the numbers given above depending on experiments being performed.

All synoptic reports from the June data sets were put through a quality control check by Aldinger (1979) to ensure a certain degree of compatability among present weather and visibility codes, in conformance with the Federal Meteorological Handbook No. 2 (U.S. Depts. of Commerce, Defense, and Transportation, 1969). All data sets including July and August 1979 data were quality-control checked by the National Climatic Center, Asheville, N.C.

D. INTERPOLATION SCHEME

All model output parameters, whose positions are within the FNOC grid, were interpolated to the ship positions from which the synoptic observations were obtained. The interpolation method used is a natural bicubic spline curvilinear scheme. This scheme and its documentation are available at the NPS W.R. Church Computer Center where all the computer computations for this study were accomplished.

E. PREDICTOR PARAMETERS

1. Model Output Parameters (MOP's)

A total of 22 analysis- and prognostic-model parameters were provided by FNOC. They were generated from the Mass

Structure Analysis model, the Primitive Equation (P.E.) model, and the Marine Wind model [U.S. Naval Weather Service, 1975]. In addition, 79 other parameters were developed from the original set. Brief descriptions of all of these parameters are listed in Appendix A.

2. Climatology Parameter

The only climatology factor used as a parameter in this study is the fog climatology developed by the National Climatic Center [Guttman, 1978]. A suitable visibility climatology was not available at the time of this study.

3. Interactive and Modified Parameters

Interactive parameters were formed in this study by using the product of two different parameters. They have been used to account for possible physical interactions between variables. Other parameters, called "modified", are simply the square, or the square root, of an MOP. A decision as to which variables to combine or modify out of an almost unlimited number of possibilities is a difficult task. Therefore, four of the parameters chosen here were taken from a previous study by Ouzts (1979). The remainder were chosen by combining or modifying those parameters which contributed significantly to explaining the variance of the predictand, in one or more experiments of this study.

4. Binary Parameters

This type of parameter is commonly used by the Techniques Development Laboratory of the National Weather

Service, Silver Springs, Maryland. A binary parameter is formed from an MOP by choosing one or more critical values of that MOP which, when equaled or exceeded, gives the binary a value of one; otherwise the binary has a value of zero. Here again, a seemingly infinite number of parameters is possible, but the set of binary parameters was limited to 14 in this study.

5. Beta Visibility Parameter

The information for the computation of this parameter was supplied by Dr. A. Goroch[3] of the Naval Environmental Prediction Research Facility. The computation uses a marine aerosol model developed for the United States Navy to test electro-optical system performance.

Apparently no formal documentation is available on the development of this model. However, Nounkester (1980) refers to this model and states that it was developed by modifying an empirical model proposed by Wells, et al., (1977). The modifications were made by B. Katz of the Naval Surface Weapons Center, White Oak, Maryland; L. Ruhnke of the Naval Research Laboratory, Washington, D.C.; and M. Munn of the Lockheed Research Laboratory, Palo Alto, California.

The aerosol model computes extinction coefficients and ranges at various wavelengths, as affected by molecular scattering and absorption, aerosol extinction and weather.

---

[3]Personal communication.

Only the visual range was of interest in this study, so only that portion of the model was used.

As input, the FNOC model output surface windspeed and relative humidity, and present weather code were supplied. Then, a parameterized visibility was computed, herein called beta visibility (BVIS). Since two relative humidity parameters were available, RHR and RHX, two beta visibility parameters could be computed, BVISR and BVISX.

Because the present weather code was not available at prognostic times, beta visibility could not be computed at tau 24 and tau 48. However, since the aerosol extinction itself was expected to correlate well with observed visibility, a modified beta visibility parameter was formed by simply omitting the weather code input. This modified beta visibility (MBVIS) could then be used at prognostic times. The method produced a less accurate parameter, but one that still correlated well with observed visibility. The methods used for computing the BVIS and MBVIS parameters are given in Appendix B.3.

IV. PROCEDURE

A. REGRESSION SCHEME

A computer program for stepwise multiple linear regression using the method of least squares was used to derive the visibility equations. The program used is one of the UCLA BMDP series, namely BMDP2R [UCLA, 1979].

In this program the dependent variable (predictand) is specified, then independent variables (predictors) are entered (forward stepping) or removed (backward stepping) based on a statistical F-test with given F-to-Enter (4.0) and F-to-remove (3.9). The first predictor selected in forward stepping is the predictor variable with the highest F-to-enter. Succeeding steps enter variables in the same manner. At each step the variables already entered into the equation are reevaluated and could be removed by backward stepping if they fail to exceed the minimum F-to-remove value.

If a variable being considered for entry reflects a strong linear combination with any of the variables already entered, it may cause computational difficulties, and the BMDP2R program will reject it if its tolerance value equals or exceeds 0.01. The program continues stepping until all variables are used, or until no further variables meet the F-to-enter value. A further definition of the statistics used is included in Appendix C.

Another regression routine available is BMDP9R, called All Possible Subsets Regression. Rather than performing a screening regression as in BMDP2R this program considers all possible combinations of predictor variables to achieve the highest possible R^2 value (explained variance). This program was used for a few experiments. Some of the computed subsets did manage to attain a higher R^2 value than that achieved by screening regression, but these R^2 values were only marginally higher and have doubtful significance. Thus, the results achieved by this method did not justify the excessive computer time involved, and so it was abandoned.

B. CATEGORICAL APPROACH

Previously at NPS, Aldinger (1979) developed analysis-time visibility regression equations based on a probability approach. Equations were developed to estimate the probability of occurrence of each of several visibility code groupings. In this study a categorical approach was used. Several schemes for grouping visibility codes into different categories were used. In order to have a visibility value for the predictand the midpoint value of the visibility range for each observed category was used. For example, if a category included synoptic codes 90-93 the visibility range would be 0-1 km, and the visibility predictand was assigned the value of 0.5 km. An exception to this rule was made for the highest visibility category. Since this category has no upper limit, several

arbitrary visibility values were assigned to the predictand depending on the categorical scheme involved. A list of the synoptic visibility codes used to determine the visibility categories can be found in the Federal Meteorological Handbook No. 2 [U.S. Depts. of Commerce, Defense and Transportation].

The regression equations so developed yield continuous visibility values (in kilometers) which can be used directly, or perhaps more appropriately, can be used to specify the selected category. The latter method is used in this study for verification purposes.

Since there are only ten reported synoptic visibility codes, with each code representing a range of visibility, the maximum number of defined categories is limited to ten. Using the maximum number of categories allows the greatest visibility resolution. However, there is some inaccuracy involved in visibility reporting that is related to an observer's ability to discriminate between different visibility ranges. Therefore, categorical schemes were developed which combined several observed codes into one category. This approach provides a wider visibility range for each category and partly compensates for observer error. It is reasoned that an observer should be able to distinguish between a few larger visibility ranges better than a larger number of smaller visibility ranges. Of course, with fewer categories some visibility resolution is lost. In the extreme case, a scheme

with only one category, which includes all visibility values, would not be affected by observer error, and all regression estimates would be perfect. However, such a scheme obviously would be useless. Therefore, some tradeoff between accuracy and resolution should be made. In this study schemes involving five and ten categories were tested.

Tau 0 equations were developed for all categorical schemes from combined June 1976 and June 1977 data. The predictor parameters considered in the equations are listed in Appendix A, part 1.

Analysis-time (Tau = 0 hr) and prognostic (Tau = 24 and 48 hr) equations were developed from July 1979 data. Prognostic equations at 24 hr and 48 hr only were developed so that the verification times would correspond to 0000 GMT. However, MOP's from 00, 12, 24, 36, and 48 hr were used. The parameter list used to develop these equations is located in Appendix A, part 2.

C. EQUATION TRUNCATION AND VERIFICATION

The BMDP2R regression routine enters a new variable at each step, increasing the R^2 value each time, thus fitting the equation better to the dependent data. After a certain number of steps, however, the incremental increase in R^2 per step may have little or no significance when the equation is applied to independent data. For this reason it was decided to truncate each equation before entering a variable which

does not increase the R^2 value by a rounded value of 1%. In general this produced an equation with four to six variables. More will be said on this topic later.

Two scoring methods were used to describe the skill of each final regression equation. These two methods consist of computing the percentage of correct forecasts and Heidke Skill score for each equation. The formula for computing these scores is given in Appendix D. The continuous visibility output from a regression equation lies within the visibility range of a particular category. This particular category is considered to be the one estimated by the regression equation. The number of times each category is thus estimated is compared to the number of observations of each category for scoring purposes.

All equations were verified against the dependent data from which they were derived. In addition, all five-category equations were verified against independent data. Equations developed from combined June 1976 and June 1977 were independently verified using July 1979 data, and equations developed from July 1979 data were verified using August 1979 data. Unfortunately, the lack of availability of MOP fields and observational data prevented the independent verification of June equations with other June data, and July equations with other July data.

Another scoring technique applies a scoring matrix developed by Aldinger (1979) and applied to the five-category

scheme. The matrix applies weights to the number of estimates of each category in order to give some credit for nearly correct estimates. This matrix, called the NPS awards matrix, is further described in Section V.C.3.

In addition, a distribution measure, called bias, is calculated for each category. Bias represents the ratio of the number of forecasts to the number of observations of each category.

V. EXPERIMENTS, RESULTS, DISCUSSION

A. CATEGORICAL SCHEMES

1. Ten-Category Scheme: 10CATA

This scheme uses ten categories of the predictand as defined below.

Category Number	Observed Visibility Code	Visibility Range (km)	Value of Predictand (km)
I	90	< 0.05	0.025
II	91	0.05 to < 0.2	0.125
III	92	0.2 to < 0.5	0.35
IV	93	0.5 to < 1.0	0.75
V	94	1.0 to < 2.0	1.5
VI	95	2.0 to < 4.0	3.0
VII	96	4.0 to <10.0	7.0
VIII	97	10.0 to <20.0	15.0
IX	98	20.0 to <50.0	35.0
X	99	≥50.0	75.0

A Tau 0 equation was developed from combined June 1976 and June 1977 data and verified on the dependent data. All values, except for regression coefficients are given to two decimal places.

Coefficient	Predictor
-354.558	
+ 1.346	EHF
+ 0.388	BVISR
+ 0.358	PS
+ 5.174	SEHF1
+ 1.380	ASTDR
- 2.938	VCMP1

$R^2 = .25$

Dependent Verification: Percent Correct = 40

Skill Score = .13

Category	I	II	III	IV	V	VI	VII	VIII	IX	X
Bias	.03	.01	.01	.01	.07	.19	.56	1.60	1.46	.01

The scores for this scheme are relatively low. The bias values indicate that the highest category and the lowest six categories are observed far more often than selected by the regression equation. On the other hand, categories VIII and IX were selected much more often than they were observed.

2. Ten-Category Scheme: 10CATB

It was felt that the arbitrarily selected midpoint value of 75.0 km for category X in 10CATA was too high, thus causing a poor fit of data in the regression equation. Therefore, this category was changed in 10CATB, as follows.

Category Number	Observed Visibility Code	Visibility Range (km)	Value of Predictand (km)
X	99	$\geq$ 50	50

All other categories, I through IX, were defined the same as in 10CATA. The Tau 0 equations was developed from combined June 1976 and June 1977 data and verified with the dependent data.

Coefficient	Predictor
-303.043	
+ 1.165	EHF
+ 0.335	BVISR
+ 0.308	PS
+ 4.627	SEHF1
+ 1.098	ASTDR
- 2.609	VCMP1

$R^2 = .28$

Dependent Verification: Percent Correct = 39

Skill Score = .13

Category	I	II	III	IV	V	VI	VII	VIII	IX	X
Bias	.03	.00	.01	.01	.05	.09	.54	1.83	1.36	.00

This equation shows some improvement over the 10CATA equation in R^2 value, however the percent correct is slightly lower and the Heidke skill score is the same.

3. Five-Category Scheme: 5CAT

Deriving a regression equation with fewer categories should yield better results due to partial compensation of observer error. In this case, five categories are used which correspond to the probabilistic five-category scheme of Aldinger (1979).

Category Number	Observed Visibility Codes	Visibility Range (km)	Value of Predictand (km)
I	90,91,92	< 0.5	0.25
II	93,94	0.5 to < 2.0	1.25
III	95,96	2.0 to <10.0	6.0
IV	97	10.0 to <20.0	15.0
V	98,99	≥20.0	35.0

The Tau 0 equation was developed from combined June 1976 and June 1977 data, and verified using both the dependent June data and independent data from July 1979.

Coefficient	Predictor
+272.710	
+ 1.035	EHF
+ 0.292	BVISR
+ 0.277	PS
+ 4.280	SEHF1
+ 0.944	ASTDR
- 0.223	VCOMP

$R^2 = .27$

Dependent Verification: Percent Correct = 44

Skill Score = .17

Category	I	II	III	IV	V
Bias	.02	.02	.47	2.12	1.05

Independent Verification: Percent Correct = 42

Skill Score = .17

Category	I	II	III	IV	V
Bias	.03	.02	.25	.87	.49

It is to be noted that the variables selected are the same as those selected in the two ten-category schemes with the exception that in this scheme VCOMP was selected instead of VCMP1. The 5CAT scheme shows an increase in skill score as expected, and the percent correct also increased. Bias values here are not much better than those for 10CATA and 10CATB except for category V of the dependent verification and category IV of the independent verification, both of which show values approaching unity.

B. REGRESSION EQUATIONS

The ultimate goal is to forecast, not just analyze, visibility. Therefore, using the July 1979 data set and a new set of parameters which included prognostic predictors, new equations were developed using the 5CAT scheme. First a new equation for Tau 0 was derived, then forecast-interval equations for Tau 24 and Tau 48 were developed. The parameter set

used for these equations is given in Appendix A, part 2. All three of the following equations were verified using the dependent data and also verified independently with data from August 1979.

1. <u>00-hr Diagnostic Equation: 5P00</u>

Coefficient	Predictor
+10.137	
+ 0.687	EHF 00
+ 0.488	BVISR
- 9.018	FTER 00
+ 3.048	SEHF1 12

$R^2 = .30$

The two-digit number after some of the predictor parameters indicates the time interval from which the parameter is derived. Those predictors without such a number are available at the analysis time only.

<u>Dependent Verification</u>: Percent Correct = 42
Skill Score = .18

Category	I	II	III	IV	V
Bias	.02	.02	.90	2.27	1.07

<u>Independent Verification</u>: Percent Correct = 51
Skill Score = .21

Category	I	II	III	IV	V
Bias	.02	.02	.99	2.00	1.10

The R^2 value and verification of equation 5P00 is better than the verification of the 5CAT equation due to the consideration of more parameters in the July 1979 data set than in the combined June 1976 and June 1979 data sets. The bias values are not much different, except for category III which shows improvement. It may be noted that all selected parameters but one are from the analysis time which seems consistent with the nature of the Tau 0 equation.

An interesting fact is that the independent verification of 5P00 yields better values than the dependent verification. This is, in part, due to the fact that the independent data contains a higher percentage of observations in those high visibility categories which the equation estimates best. In addition the dependent data comes from a large enough sample of synoptic conditions that the regression equation could score higher when applied to independent data, which by chance includes a larger number of those synoptic situations best handled by the equation.

2. <u>24-hr Prognostic Equation: 5P24</u>

Coefficient	Predictor
+ 0.085	
+ 1.077	EHF 24
+ 0.440	BVISR
+ 0.002	RHRX
- 7.418	FTER 24

$R^2 = .30$

Dependent Verification: Percent Correct = 42

Skill Score = .16

Category	I	II	III	IV	V
Bias	.10	.08	.56	2.26	1.16

Independent Verification: Percent Correct = .52

Skill Score = .20

Category	I	II	III	IV	V
Bias	.04	.07	.61	1.92	1.17

There is a deterioration in R^2 value when 5P24 is compared to 5P00, as one might expect. The percent correct is similar for both equations, but the Heidke skill score for 5P24 is slightly less than for 5P00. Here again, as in 5P00, the independent verification is better than the dependent verification.

It is to be noted that variables from Tau 24 have entered the 5P24 equation, which is consistent with the nature of a Tau 24 equation.

3. 48-hr Prognostic Equation: 4P48

Coefficient	Predictor
- 4.160	
+ 0.390	EHF 36
+ 0.555	BVISR
-12.631	FTER 48
+ 0.633	EHF 00
+ 0.003	RHRSQ
- 0.160	MBVIS 48

$R^2 = .27$

Dependent Verification: Percent Correct = 42

Skill Score = .13

Category	I	II	III	IV	V
Bias	.01	.01	.29	2.08	1.40

Independent Verification: Percent Correct = 52

Skill Score = .16

Category	I	II	III	IV	V
Bias	.00	.01	.20	1.72	1.32

Here the R^2 value has deteriorated somewhat from the 5P00 and 5P24 cases. The percent correct is the same for equations at all three time periods, but the Heidke skill score in 5P48 is worse than that for 5P24 and 5P00. Overall the bias values for 5P48 are worse than for both 5P00 and 5P24. Once again the independent verification is better than the dependent verification.

It is to be noted that two Tau 48 hr predictors have entered the equation. However, there is also one TAu 36 hr predictor and three Tau 00 hr predictors. The predictor BVISR shows up in 5P48 as well as in 5P00 and 5P24. BVISR, which itself is a parameterized visibility, can be considered an indicator of the persistence of marine visbility regimes through 48 hours.

C. PROBABILISTIC VS. CATEGORICAL APPROACH

Aldinger (1979) used the 5CAT scheme outline previously and developed regression equations for the probability of

occurrence of each category. Then, using the notion of threshold probability, the most-likely category was determined. For comparison, an equation was developed by the categorical method of this study considering only those predictor parameters used by Aldinger. All equations were derived from the combined June 1976 and June 1977 data and were verified dependently.

1. <u>Probabilistic Equations [Aldinger, 1979]</u>

Category	Equation
I	VISPROB = 366.262 - 1.647 SEHF + .289 RHR - .369 PS + .401 VCOMP
	$R^2 = .13$
II	VISPROB = 738.837 - .264 EHF - .746 PS + .555 RHR - 1.689 SEHF
	$R^2 = .21$
III	VISPROB = 266.075 + .303 VVWW - .256 PS + .247 RHR + .313 RHX
	$R^2 = .05$
IV	VISPROB = -278.669 + .365 SEHF - .643 VCOMP + .431 VVWW + .333 PS
	$R^2 = .09$
V	VISPROB = -693.510 + 3.633 EHF + .767 PS - .709 VCOMP - .352 RHR
	$R^2 = .21$

VISPROB is the probability of occurrence of the category for which the equation is derived.

Dependent Verification: Percent Correct = 32

Skill Score = .13

Category	I	II	III	IV	V
Bias	.04	1.53	1.10	2.08	0.40

2. Categorical Equation

Only one categorical equation was derived whose visibility value (VIS) determines the visibility category by selecting that category to which VIS belongs.

$$VIS = -302.35 + .175\ EHF + .339\ PS - .254\ RHR + .730\ SEHF$$

$$R^2 = .24$$

Dependent Verification: Percent Correct = 43

Skill Score = .14

Category	I	II	III	IV	V
Bias	.02	.01	.28	2.08	1.13

Comparing the two approaches shows that the categorical approach yields a higher percent correct and a slightly higher skill score. However, except for category V, the biases are worse for the categorical scheme. As might be expected both methods use similar predictor parameters. SEHF, RHR, PS and EHF are common to both.

3. NPS Awards Matrix

Aldinger (1979) developed an awards matrix which when applied to the verification matrix (Appendix E) of a

5-category scheme gives some credit to near successes. The Techniques Development Laboratory (TDL) of the National Weather Service has also used an awards matrix, but of a different nature, which does not give full credit to all correct visibility estimates [National Weather Service, 1973]. The NPS awards matrix does give full credit to all correct estimates. All quantities of a verification matrix are multiplied by the corresponding percentages in the awards matrix shown below.

OBSERVED	Estimated Category				
CATEGORY	I	II	III	IV	V
I	100	80	0	0	0
II	80	100	25	0	0
III	0	25	100	25	0
IV	0	0	25	100	75
V	0	0	0	75	100

The verification results, after applying the awards matrix, are as follows:

<u>Probabilistic Approach</u>: Percent Correct = 60

Skill Score = .27

<u>Categorical Approach</u>: Percent Correct = 63

Skill Score = .12

In both cases percent correct increases markedly. However, for the probabilistic approach the skill score doubles,

while for the categorical approach the skill score decreases. This shows that the probabilistic approach forecasts near successes much better than the categorical approach, thus enhancing its usefulness.

D. PREDICTAND TRANSFORMATIONS

Generally the relationship between an atmospheric predictand and the predictors is not linear. This can lead to less than desirable results when multiple linear regression is used. Non-linear regression may be used to overcome this problem, but the increased computational time involved usually precludes its use. Another method used to solve the non-linear problem is to transform the predictand to a form which then relates in a more linear manner to the predictors.

Using a limited number of parameters several transforms were tested on the 10CATA scheme, using July 1976 and July 1977 data. The relative values of R^2 produced using each transform are shown below.

Predictand	R^2
VISIBILITY (VIS)	.230
Log_{10}(VIS)	.243
1/VIS	.037
$(1/\text{VIS})^2$	.011
$\text{VIS}^{1/2}$	.272
$\text{VIS}^{1/3}$	.273
$\text{VIS}^{1/4}$	.267

It can be seen that the R^2 value for several of the transformed predictands was higher than the R^2 value for the non-transformed visibility predictand, though the increase was not large.

However, the real test is how well an equation with a transformed predictand verifies. So the equation derived with the cube root of visibility as the predictand, which yielded the highest R^2 value, was scored against the equation with the non-transformed predictand.

<u>Predictand = visibility</u>.

<u>Dependent Verification</u>: Percent Correct = 39

Skill Score = .14

<u>Predictand = visibility$^{1/3}$</u>

<u>Dependent Verification</u>: Percent Correct = 27

Skill Score = -.01

The results show that the transformed predictand yielded worse scores than the unmodified visibility predictand. This is a surprising result in view of the relative R^2 value. It may, in part, be explained by the fact that there was an uneven distribution of visibility observations between categories, with a heavy weighting toward higher visibility categories. Time limitations, however, did not permit examining this further, and all other research was conducted using the non-transformed predictand.

E. WEIGHTED LEAST SQUARES

In this study the data distribution is such that most observations occurred in the higher categories, in particular category 98. The result of this is a regression equation that fits the higher visibility categories better than the lower visibility categories. As a result, low visibilities are poorly estimated.

The technique of weighted least squares was applied in an attempt to alleviate this problem. The goal was to weight more heavily the lower category cases in relation to those in the higher categories so that the resultant equation would increase skill in estimating poor visibilities.

The BMDP programs [UCLA, 1979] allow case weights to be applied. The weighted least squares technique minimizes

$$w_j \sum (y_j - \hat{y}_j)^2$$

where,

w_j is the case weight for case j

y_j is the observed visibility for case j

$\hat{y}_j$ is the regression estimate for case j.

Normally the weight for each case should be inversely proportional to the variance [Daniel, 1971], but any number of weighting techniques may be tried. In this study two sets of case weights were tried and applied to the schem of 10CATA.

The first scheme (WLS1) weighted each case with a weight equal to the inverse of the predictand value, as follows.

For cases of observed code	The predictand value (km) is	And the case weight (w_j) is
90	.025	1/.025
91	.125	1/.125
92	.35	1/.35
93	.75	1/.75
94	1.5	1/1.5
95	3.0	1/3.0
96	7.0	1/7.0
97	15.0	1/15.0
98	35.0	1/35.0
99	75.0	1/75.0

The resultant equation derived from combined June 1976 and June 1977 data (not given here) was verified dependently with the following results.

R^2 = .09

Percent Correct = 7

Skill Score = -.01

Obviously, this is a poor weighting system. The R^2 value is very low and the scores are predictably poor.

For the second scheme (WLS2) a more reasonable set of weights was used. The variance was computed for each category from the unweighted equation of 10CATA. Then the weight

for each case in a particular observed category was set to the inverse of the square root of the variance of the observed category.

For Cases of Observed Code	The Predictand value (km) is	And the case weight (w_j) is
90	.025	.0052
91	.125	.0603
92	.35	.0661
93	.75	.0615
94	1.5	.0702
95	3.0	.0700
96	7.0	.0754
97	15.0	.0941
98	35.0	.0925
99	75.0	.0242

(Each code group corresponds to a category in the 10CATA scheme.)

The case weights shown here are somewhat contrary to what might be expected. It would seem that the variances of the higher categories would be larger than those of the smaller categories, if for no other reason than the fact that the visibility ranges of the higher categories are greater. If this were true the case weights for the higher categories would be smaller than for the lower categories. However, the weights shown here generally increase with an increase in

category, with the exception of category X (code 99). This result is due to the fact that the regression equation estimates those categories best which contain the highest number of observations, namely the categories containing codes 97 and 98.

A comparison of dependent verification between the equations of 10CATA and WLS2 shows very little difference.

Scheme	R^2	Percent Correct	Skill Score
10CATA	.25	40	.13
WLS2	.23	40	.12

F. DEFLATION OF R^2

According to theory, if a regression equation perfectly fits the data from which it was developed the explained variance, R^2, should equal 1.0. However, it appears that due to the nature of the categorical schemes in this study a limit was placed on the maximum R^2 that it was possible to achieve. This particular limit is related to the fact that each predictand value was assumed to be the midpoint value of the observed category, thus providing discrete visibility values. However, the regression equation gives continuous visibility values which are then used with the assigned predictand values to determine R^2.

In one experiment, to demonstrate the deflation of R^2, a regression equation of the form of 10CATA scheme was developed. Then using the dependent data, the equation was

used to compute visibility values, V_i.

Symbolically: $V_i = A_1 + B_1 x_{1i} + C_1 x_{2i} + \ldots$

where,

V = visibility

x's = independent predictors.

These V_i values were used as substitutes for the original visibility observations. Next, using these V_i values, a new predictand, V_i', was derived by re-setting the V_i value to the midpoint of the category to which V_i belonged, giving V_i'. Finally, a second regression equation was developed using the V_i' as predictand values to yield an equation of the form

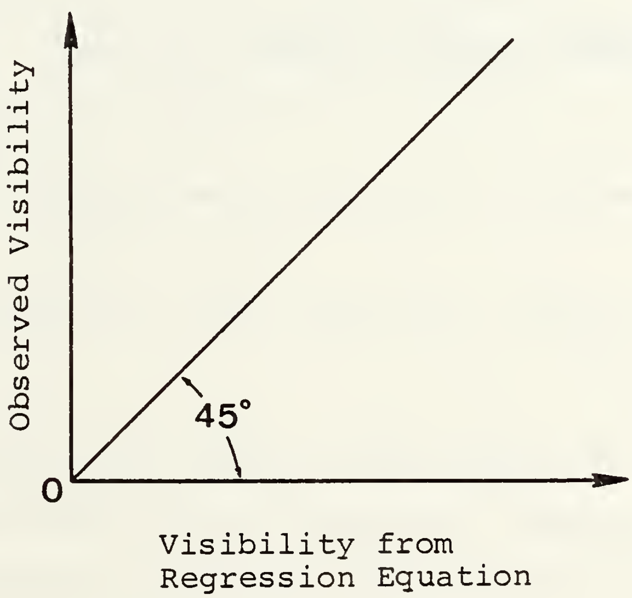
$$V_i'' = A_2 + B_2 x_{1i} + C_2 x_{2i} + \ldots .$$

It can be seen that if the continuous values, V_i, had been used as the predictand the second regression equation would be identical to the first one and have an R^2 value of 1.0. However, because the predictand, V_i', used to develop the second equation has discrete values as defined by the categorical scheme, the second equation is not identical to the first; and the R^2 value is approximately 0.7, using V_i' as the observed values.

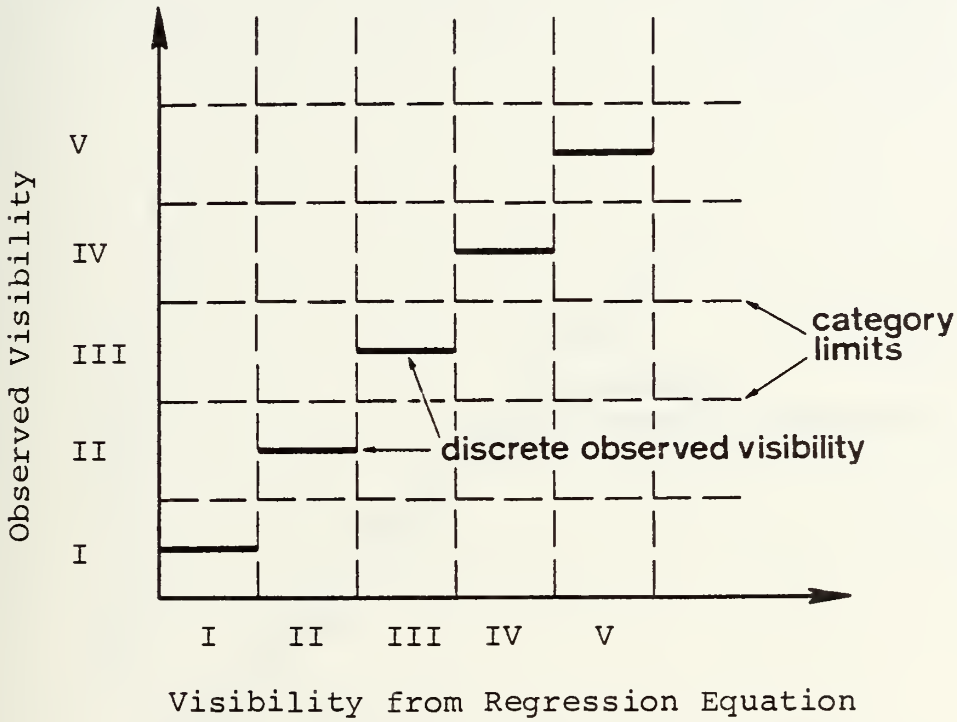
It is believed that the R^2 value of 0.7 rather than 1.0 is the maximum value achievable in the 10CATA scheme with a

perfect equation, due to the method of defining the predictand used in this study. The other categorical schemes, of course, have a similar R^2 limit.

The drop of R^2 from 1.0 to 0.7 can be demonstrated by schematic graphs. Assuming that the observed visibility can be expressed perfectly by a regression equation, for which $R^2 = 1.0$, then the graph below is the result. As the continuous regression-estimated visibility increases the observed visibility increases continuously also.



However, the observed visibility is not given as a continuous variable. Rather the visibility observations are given as ranges or categories, and the visibility predictand is defined as the midpoint of the observed range, which is demonstrated schematically below.

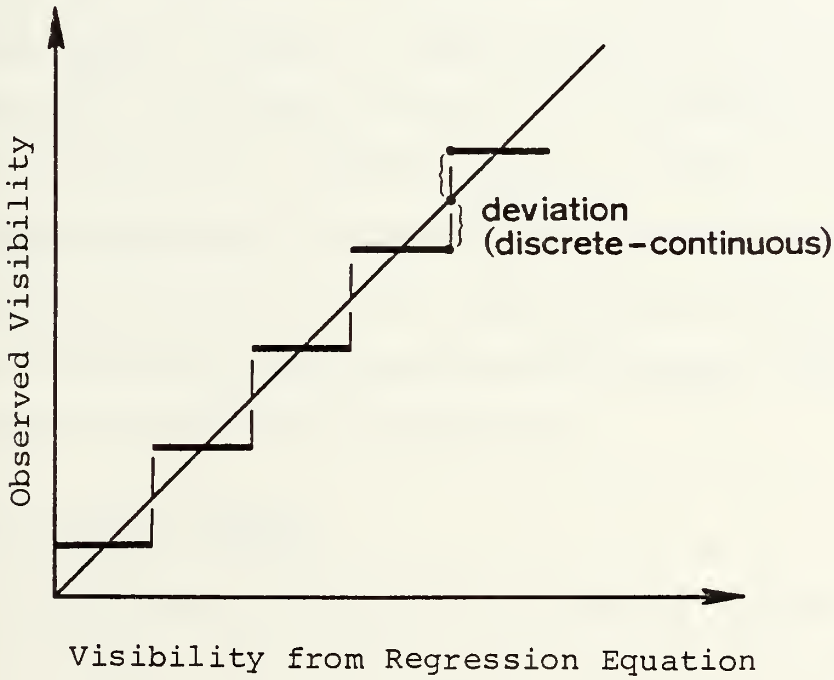


The schematic above shows a step function relationship which indicates that as the continuous regression-estimated visibility increases within each categorical visibility range (given by roman numerals) the observed visibility remains constant.

The regression-estimated visibility values have not changed from the first schematic to the second but the

verifying "observed" values have changed from continuous to discrete values. All observed values below a categorical midpoint value have been increased, and values lying above a midpoint value have been decreased.

The deterioration of R^2 which results from the second case can be seen by noting the deviation of values along the discrete observed visibility step function from the continuous observed visibility line as shown below.



In another experiment, an attempt was made to compute the R^2 value for the 10CATA equation without the hindrance of the problem just described. The BMDP programs compute R^2 using the continuous regression-produced visibility

values and the discrete observed values. A separate program was developed to compute R^2 by first re-setting the continuous regression values of 10 CATA to the midpoint values of the categories to which they belong. Then, using the discrete predictand values, a new R^2 was computed. In this case discrete values are used for both the observations and the regression estimates. The R^2 value computed in this way is .31 as compared to .25 computed by the BMDP programs. All R^2 values previously shown in this study were computed by the method used in the BMDP programs.

The maximum R^2 value of approximately 0.7 as found by experiment for the 10CATA scheme may be compared to the R^2 value of .31 which the 10CATA equation yielded. The difference between the two R^2 values of approximately 40% can now be attributed to errors in the observations and numerical MOP's and the non-linear relationship between visibility and associated meteorological parameters.

G. DISTRIBUTION PROBLEM

The distribution of observations among synoptic codes for the combined June 1976 and June 1977 data set is shown below. It can be noted that the highest three categories contain 66% of the observations, and the highest four categories contain 79% of the observations. The observation distributions are similar for the July 1979 and August 1979 data sets.

Code Group	Number of Observations	Percent of Total observations
90	75	0.8
91	238	2.6
92	400	4.4
93	740	8.1
94	166	1.8
95	327	3.6
96	1125	12.3
97	1911	21.0
98	3642	39.9
99	495	5.4

This fact tended to tune all the regression equations to the high categories, such that high categories were estimated relatively well by the regression equations and low visibility categories were estimated poorly. This is somewhat contrary to what is desired, since forecasts of low visibility are very important operationally.

The probabilistic approach does not have a similar distribution problem, since one regression equation is developed for each visibility category and depends only on the observations of a single category.

H. BETA VISIBILITY

The beta visibility was previously described. Its computation is given in Appendix B.3. Beta visibility is not only

a parameter for use in visibility regression equations but itself yields a value of visibility which may be of use. This section attempts to quantify its usefulness.

The BMDP programs were used to compute a correlation coefficient between the predictand and the various forms of the beta visibility parameter. It is to be noted that the visibility predictand is not a directly observed visibility value, but rather it is the midpoint value of an observed visibility range. The correlation coefficients, R, between the various forms of the beta visibility parameter and the visibility predictand of the 5CAT scheme are given in the following table. A comparison of maximum, minimum and mean values is also given. These statistics were derived using the July 1979 data set.

Comparative Statistics and Correlation to the Visibility Predictand (VIS) at Tau 0 hr

	Maximum (km)	Minimum (km)	Mean (km)	R
VIS (Tau 0)	35.0	0.25	19.2	1.00
BVISR	46.9	0.56	14.3	0.43
BVISX	51.9	0.79	19.9	0.09

Comparative Statistics and Correlation to the Visibility Predictand (VIS) at Tau 0+24 hr

	Maximum (km)	Minimum (km)	Mean (km)	R
VIS (Tau 24)	35.0	0.25	19.0	1.00
BVISR	48.7	0.51	14.3	0.31
BVISX	51.9	0.79	20.0	0.10
MBVIS 24	44.4	1.68	17.2	0.05

Comparative Statistics and Correlation to the Visibility Predictand (VIS) at Tau 0+48 hr

	Maximum (km)	Minimum (km)	Mean (km)	R
VIS (Tau 48)	35.0	0.25	18.8	1.00
BVISR	52.1	0.42	14.3	0.24
BVISX	51.9	0.62	20.0	0.06
MBVIS 48	50.1	2.14	15.4	0.02

It should be noted that in the table the analysis-time parameters BVISR and BVISX are compared to the predictand at all three time periods. The table shows that the maximum, minimum and mean values of all the beta visibility parameters are similar to the corresponding values of the visibility predictand at each time period. BVISR shows a higher correlation to the predictand than BVISX at all time periods, though the correlation of both parameters to the predictand worsens with time. Both the analysis-time parameters BVISR and BVISX show higher correlation to the predictand at Tau 24 hr than the prognostic-time parameter MBVIS 24. The same is true at Tau 48 hr when comparing BVISR and BVISX to MBVIS 48.

The following clarifies the reason for the slight differences in maximum, minimum and mean values for the same parameter at different time periods. The Tau 24 hr data includes values from the first day of August (i.e. up to 24 hrs after the last day of the July data set), and

omits values from the first day in July. In like manner, the Tau 48 hr data includes the first two days of August and omits the first two days of July. Thus the data set for each time period is slightly different.

In addition, a skill score was computed for BVISR and BVISX by determining the code group to which the computed beta visibility belonged, and comparing that to the observed code groups in the combined June 1976 and June 1977 data.

	Heidke Skill Score	Percent Correct
BVISR	0.10	33
BVISX	0.07	31

It can be concluded by these results that although beta visibility is a useful predictor parameter for regression analysis, it has quite limited skill when used to estimate visibility by itself.

I. COMMENTS ON EXPLAINED VARIANCE

The total explained variance, R^2, of a multiple linear regression equation is a measure of how well the dependent variable (predictand) can be approximated by a linear combination of independent variables (predictors). The higher the value of R^2, the better the approximation is. A perfect linear relationship results in an R^2 value of 1.0. However, it should be noted that R^2 indicates only how well a given equation will estimate a given predictand if one uses the

method of least squares. This method results in a regression equation which minimizes the value of the sum of squares of the estimate errors (estimate error = estimated value minus observed value). An equation with a given R^2 will not necessarily provide a better estimate of the predictand than an equation with a smaller R^2 when evaluated by some method other than least squares. An entirely different situation may occur if one applies the derived regression equation to independent data. Though the original equation may be a good fitting equation for the dependent data (by the least squares criterion) it may be a poor fit for the independent data, especially if the number of cases is small. In this study the sample size of over 4000 cases is large enough that a drastic drop in estimation ability is not to be expected when independent data are applied, however some deterioration was encountered.

Also, as additional predictors are entered into an equation by the stepwise process the R^2 value will increase, but an equation with fewer predictors and a lower R^2 may, in fact, provide a better estimate when applied to independent data. This is so, since as more variables enter into an equation, it becomes more likely that the equation will reflect relationships unique to the dependent data. Thus extra variables may degrade an equation when scored on independent data [Air Weather Service, 1977]. Of course, the application of independent data may also show an improvement in scores due to

the peculiarities of a particular data set. However, some form of truncation method should be used to limit the number of variables in an equation such as was done in this study.

An experiment to demonstrate the relationship of score to number of predictors in the equation was performed, using the regression results of the 5CAT scheme. Truncating the 5CAT scheme at different steps yielded the following.

	Dependent Data			Independent Data	
Step	R^2	Skill Score	% Correct	Skill Score	% Correct
1	.166	.123	40.4	.128	39.5
2	.219	.149	42.7	.173	41.8
3	.245	.153	44.0	.179	42.7
4	.256	.151	43.2	.178	43.2
5	.262	.167	43.8	.179	42.7
6	.269	.174	44.0	.165	41.9
7	.272	.166	44.4	.156	41.2
8	.275	.174	44.0	.163	40.9

It can be seen that after a certain point the direct relationship between R^2 and skill becomes obscure. In this study the equation for the 5CAT scheme as described in the text was truncated after the sixth step, for at the seventh step the R^2 failed to increase by a rounded value of 1%.

It is encouraging to note that the results above show that percent correct and skill score do not substantially decrease when independent data is applied compared to when dependent data is applied. In fact, the skill score is relatively better in the former instance for the first five steps.

J. DISCUSSION OF ERRORS

It is believed by the author that the techniques used in this study would yield equations of high operational usefulness if it were not for various unavoidable errors. Linear regression assumes, for example, that all predictand values used are errorless. This is far from true here. Observer error in estimating visibility at sea is relatively high, due mostly to a dearth of visibility markers at sea and also due to the fact that many ships transmitting synoptic reports may have observers with little or no observational training and/or experience.

Errors also enter into the Model Output Parameters, which are only as good as the numerical models from which they are generated, analyses being better than prognosis. The method used to interpolate the MOP's to the synoptic ship positions also adds error to the scheme.

VI. CONCLUSIONS AND RECOMMENDATIONS

The categorical approach used in this study yielded visibility equations which have comparable skill both at analysis and prognostic times which is a promising result. However, the actual skill of the equations is relatively poor and not operationally useful at this time. The reason for this is believed to lie inherent in the errors of visibility observations, the non-linear relationship between the predictand and the predictors, and the numerically generated MOP's. The future promises much improvement due to new statistical techniques, improved numerical models and the identification of more air/ ocean parameters with a known relation to visibility.

The comparison of the probabilistic to the categorical approach indicates that the probabilistic approach holds more promise, at least partly due to the fact that the categorical approach is hindered by the uneven distribution of observations. The probabilistic approach seems to estimate near successes better than the categorical approach.

Parameters found to be most highly related to visibility in the regression equations are: evaporative heat flux, beta visibility, sea level pressure, sensible plus evaporative heat flux, air/sea temperature difference,

meridional component of the wind, relative humidity parameters and FNOC's fog probability parameter.

The following recommendations are offered for future research:

1. Test new parameters in relation to visibility, such as some type of visibility persistence parameter, more interactive, modified and binary parameters, and a climatological parameter now being developed for the North Pacific by the National Climatic Center.

2. Investigate further the techniques of weighted least squares and transformation of the predictand to relate more closely to the non-linear nature of the problem.

3. Stratify the data with respect to critical values of geography and to various MOP's.

4. Investigate the use of discriminant analysis to estimate visibility.

5. Stress the probabilistic approach over the categorical approach, and in particular, expand the work of Aldinger [1979] to include additional parameters and prognostic equations.

APPENDIX A

PREDICTOR PARAMETER DESCRIPTIONS

Part 1. This part consists of all predictor parameters considered for use in the analysis-time equations developed from the combined June 1976 and June 1977 data set.

NOTES:

[**] Denotes those predictor parameters that repeatedly were selected early by the stepwise regression thereby implying their relatively strong relationship with visibility.

[*] Denotes those predictor parameters that only occasionally or never were selected early by the stepwise regression, but may be useful in future studies.

[-] Denotes those predictor paramters that seemed to have little or no relation to visibility in this study.

SYMBOL	DESCRIPTIVE NAME	UNITS
A.	Analysis Parameters (FNOC Mass Structure Model)	
PS	Sea-level Pressure [**]	(mb)
TAIR	Surface Air Temperature [*]	(°C)
EAIR	Surface Vapor Pressure [*]	(mb)
T925	925 mb Air Temperature [*]	(°C)
TSEA	Sea-Surface Temperature [*]	(°C)

B. Prognostic Parameters (FNOC Primitive Equation Model)

TX	Surface Air Temperature [*] Derived from surface air and potential temperatures, boundary layer depth, upper-level winds extrapolated to surface, air density, drag coefficient, gustiness factor and empirical constants.	(°C)
EX	Surface Vapor Pressure [*] Derived from model's mixing ratio	(mb)
SOLARAD	Solar Radiation [*] Calculated absorption of incoming short-wave (solar) radiation. (postive downward)	(gcal/ cm^2/hr)
EHF	Evaporative Heat Flux [**] Derived using air density, drag coefficient extrapolated winds, and mixing ratios.	(gcal/ cm^2/hr)
SHF	Sensible Heat Flux [*] Recovered from SHF = SEHF-EHF. Originally derived by FNOC using drag coefficient, extrapolated winds, surface air temperature, TX, density and constants.	(gcal/ cm^2/hr)
SEHF	Sensible Plus Evaporative Heat Flux [**] SEHF = SHF+EHF	(gcal/ cm^2/hr)
THF	Total Heat Flux [*] THF = SEHF-SOLARAD+LW, where LW is the heating due to long-wave (terrestrial) radiation.	(gcal/ cm^2/hr)

C. Marine Wind Model (FNOC)

VVWW	Marine Wind Speed [*]	(kt)
(DDWW)	Marine Wind Direction This variable was not used as a predictor parameter, but rather to derive other parameters.	(deg/10)

D. Derived Parameters

UCOMP	Zonal Wind Component [*] UCOMP = -VVWW sin(DDWW·10)	(m/sec)

VCOMP	Meridional Wind Component [**] VCOMP = -VVWW cos(DDWW·10)	(m/sec)
CAPU	I Directional Wind Component [*] CAPU = -UCOMP · sin(LNGA) -VCOMP · cos(LNGA) [Haltiner, 1971], where LNGA = -10 - (I,J point longitude).	(m/sec)
CAPV	J Directional Wind Component [*] CAPV = VCOMP · cos(LNGA) -VCOMP · sin(LNGA) [Haltiner, 1971], where LNGA = -10 - (I,J point longitude).	(m/sec)
THETAX	Potential Temperature X [-] Derived using PS, TX.	(°K)
THETAR	Potential Temperature R [-] Derived using PS, TAIR.	(°K)
STABX	Stability X [-] Derived using [THETAX - (THETA from T925)]/(PS-925].	(°K/mb)
STABR	Stability R [-] Derived using [THETAR - (THETA from T925)]/(PS-925).	(°K/mb)
ASTDX	Air-Sea Temperature Difference X [**] ASTDX = TX-TSEA	(°C)
ASTDR	Air-Sea Temperature Difference R [**] ASTDR = TAIR-TSEA.	(°C)
ADTSEA	Advection of TSEA [*] See Appendix B.1.	(°C/hr)
ADTX	Advection of TX [*] See Appendix B.1.	(°C/hr)
ADTAIR	Advection of TAIR [-] See Appendix B.1.	(°C/hr)
AASTDX	Advection of ASTDX [-] See Appendix B.1.	(°C/hr)
AASTDR	Advection of ASTDR [*] See Appendix B.1.	(°C/hr)

RHR	Relative Humidity R [**] See Appendix B.2.	(%)
RHX	Relative Humidity X [**] See Appendix B.2.	(%)

E. Interactive and Modified Parameters

RHRX = RHR · RHX [**]

RVCOMP = RHR · VCOMP [-]

RHRPS = RHR · PS [-]

RASTDX = RHR · ASTDX [**]

RSEHF = RHR · SEHF [-]

PDSQ = $(PS-1014.8)^2$ [-]

PSRHX = PS · RHX [-]

PSSEHF = PS · SEHF [-]

PASTDX = PS · ASTDX [*]

PSVCMP = PS · VCOMP [-]

VSEHF = VCOMP · SEHF [-]

EHFADT = EHF · ADTAIR [-]

ESEHF = EHF · SEHF

EXEAIR = EX · EAIR [-]

SEVCMP = SEHF · VCOMP [-]

SEADTX = SEHF · ASTDX [-]

SERHX = SEHF · RHX [-]

ASTDRX = ASTDR · ASTDX [*]

UVCOMP = UCOMP · VCOMP [*]

CAPUV = CAPU · CAPV [*]

TARSEA = TAIR · TSEA [-]

TXAIR = TX · TAIR [-]

SEHFSQ	= SEHF · SEHF [-]
EHFSQ	= EHF · EHF [-]
RHRSQ	= RHR · RHR [**]
RHXSQ	= RHX · RHX [*]
VCMPSQ	= VCOMP · VCOMP [-]
CAPUSQ	= CAPU · CAPU [*]
TSEASQ	= TSEA · TSEA [-]
ASDXSQ	= ASTDX · ASTDX [**]
ASDRSQ	= ASTDR · ASTDR [*]
ADSESQ	= ADTSEA · ADTSEA [-]
PSSQ	= PS · PS [-]
SREHF	Square root of EHF [*]
SRPS	Square root of PS [*]
SRASTR	Square root of ASTDR [-]
SRASTX	Square root of ASTDRX [-]
SRSEHF	Square root of SEHF [*]
SRRHR	Square root of RHR [-]
SRRHX	Square root of RHX [-]
SRCAPU	Square root of CAPU [-]
SRTSEA	Square root of TSEA [-]
SRVCMP	Square root of VCOMP [-]
SRASEA	Square root of ADTSEA [*]

F. Binary Parameters

EHF1 $\begin{cases} \text{if } EHF < 1.75 \text{ or } EHF > 8.75;\ EHF1 = 0.0\ [-] \\ \text{if } 1.75 \leq EHF \leq 8.75;\ EHF1 = 1.0 \end{cases}$

EHF2 $\begin{cases} \text{if EHF} < 3.34;\ \text{EHF2} = 0.0 \\ \text{if EHF} \geq 3.34;\ \text{EHF2} = 1.0 \end{cases}$ [*]

EHF3 $\begin{cases} \text{if EHF} < 0.0;\ \text{EHF3} = 0.0 \\ \text{if EHF} \geq 0.0;\ \text{EHF3} = 1.0 \end{cases}$ [-]

PS1 $\begin{cases} \text{if PS} < 1000 \text{ or PS} > 1030;\ \text{PS1} = 0.0 \\ \text{if } 1000 \leq \text{PS} \leq 1030;\ \text{PS1} = 1.0 \end{cases}$ [-]

PS2 $\begin{cases} \text{if PS} < 1014.8;\ \text{PS2} = 0.0 \\ \text{if PS} \geq 1014.8;\ \text{PS2} = 1.0 \end{cases}$ [-]

RHR1 $\begin{cases} \text{if RHR} < 60;\ \text{RHR1} = 0.0 \\ \text{if RHR} \geq 60;\ \text{RHR1} = 1.0 \end{cases}$ [-]

RHR2 $\begin{cases} \text{if RHR} < 83;\ \text{RHR2} = 0.0 \\ \text{if RHR} \geq 83;\ \text{RHR2} = 1.0 \end{cases}$ [-]

SEHF1 $\begin{cases} \text{if SEHF} < 0.0;\ \text{SEHF1} = 0.0 \\ \text{if SEHF} \geq 0.0;\ \text{SEHF1} = 1.0 \end{cases}$ [**]

ASDX1 $\begin{cases} \text{if ASTDX} < 0.0;\ \text{ASDX1} = 0.0 \\ \text{if ASTDX} \geq 0.0;\ \text{ASDX1} = 1.0 \end{cases}$ [-]

ASDR1 $\begin{cases} \text{if ASTDR} < 0.0;\ \text{ASDR1} = 0.0 \\ \text{if ASTDR} \geq 0.0;\ \text{ASDR1} = 1.0 \end{cases}$ [-]

VCMP1 $\begin{cases} \text{if VCOMP} < 0.0;\ \text{VCMP1} = 0.0 \\ \text{if VCOMP} \geq 0.0;\ \text{VCMP1} = 1.0 \end{cases}$ [**]

UCMP1 $\begin{cases} \text{if UCOMP} < 0.0;\ \text{UCMP1} = 0.0 \\ \text{if UCOMP} \geq 0.0;\ \text{UCMP1} = 1.0 \end{cases}$ [-]

STABX1 $\begin{cases} \text{if STABX} < 0.0;\ \text{STABX1} = 0.0 \\ \text{if STABX} \geq 0.0;\ \text{STABX1} = 1.0 \end{cases}$ [-]

STABR1 $\begin{cases} \text{if STABR} < 0.0;\ \text{STABR1} = 0.0 \\ \text{if STABR} \geq 0.0;\ \text{STABR1} = 1.0 \end{cases}$ [-]

G. Other Parameters

FTER	FNOC Fog Probability Parameter [**]	(%)
BVISR	Beta Visibility Parameter R [**] See Appendix B,3.	(km)
BVISX	Beta Visibility Parameter X [*] See Appendix B,3.	(km)

Part 2. This part consists of all predictor parameters considered for use in the analysis-time and forecast-interval equations developed from the July 1979 data. In this list some parameters not found useful in the June regression runs were eliminated, but additional parameters which were available for the July data set were added.

A. Predictors used to develop equations both from June and from July data (described in Part 1)

(1) Parameters available for Tau 00, 12, 24, 36 and 48 hr

PS	T925	TX
EX	EHF	SHF
SEHF	THF	VVWW
UCOMP	VCOMP	RHX
EHF2	SEHF1	VCMP1
FTER		UVCOMP

(2) Parameters available for Tau 00 hr only

TAIR	EAIR	TSEA
ASTDX	ASTDR	RHR
ASTDRX	ASDXSQ	RASTDX
RHRX	RHRSQ	BVISR
BVISX		

B. Additional variables available in the July 1979 data set

SYMBOL	DESCRIPTIVE NAME	UNITS
CLIMO	National Climatic Center Fog Frequency Climatology [*]	(%/100)
SSANOM	Sea Surface Temperature Anomaly [*] Available at Tau 00 hr	(°C)
U925	U Wind component at 925 mb [*] Available at Tau 00, 12, 24, 36, 48 hr	(kt)
V925	V Wind component at 925 mb [*] Available at Tau 00, 12, 24, 36, 48 hr	(kt)
E925	Vapor pressure at 925 mb [*] Available at Tau 12,24,36,48 hr	(mb)
GGTHTA	Front Location Parameter [*] Available at Tau 00, 12, 24, 36, 48 hr	$(°K/(100\ km)^2)$
NCLOUD	Total Cloud Cover [*] Available at Tau 00, 12, 24, 36, 48 hr	(tenths)
MBVIS	Modified beta visibility [**] See Appendix B.3 Available at Tau 12, 24, 36, 48 hr	(km)
RASTDR	= RHR · ASTDR [*] Available at Tau 00 hr	(°C %)
H510	1000 mb - 500 mb [*] D-value thickness Available at Tau 00, 12, 24, 36, 48 hr	(cm)

APPENDIX B

MISCELLANEOUS PARAMETER FORMULATIONS

1. Advection Parameters

All advection parameters use the following general formulation.

For the advection of a quantity (Q) the formula $ADQ = -\vec{V} \cdot \nabla Q$, was used in the finite difference form:

$$ADQ = -\frac{RMAP}{DM}\,[CAPU\cdot(Q_{I+1} - Q_{I-1})_J + CAPV\cdot(Q_{J+1} - Q_{J-1})_I],$$

where RMAP = (1 + sin 60)/(1 + sin (latitude))

and DM = $[2\cdot(6.37\cdot 10^6)\cdot(1 + \sin 60)]/31.205$

(31.205 = number of grid mesh lengths, pole to equator, on the FNOC I,J grid).

2. Relative Humidity Parameters

The thermodynamic equation for calculation of saturation vapor pressure, known as the Clausius-Clapeyron equation is given as

$$\frac{1}{e_s}\frac{de_s}{dT} = L(T)/RT^2 \quad . \qquad (1)$$

where

R = specific gas constant for water vapor

(0.461 joule g^{-1} $°K^{-1}$)

T = temperature (°K)

$L(T)$ = latent heat of vaporization of water (joule g^{-1})

e_s = saturation vapor pressure.

This describes the behavior of e_s as a function of T, assuming water vapor to be an ideal gas. It cannot be integrated exactly to give e_s as a function of T, since L(T) is not known to sufficient accuracy at more than a few temperatures [Weinreb, 1971].

The Goff/Gratch formula (Eq. 2) is an approximate solution of Eq. (1) considering the deviations from a perfect gas based on modern experimental data [List, 1963].

$$\begin{aligned} \log_{10} e_s = & -7.90298(T_s/T-1) + 5.02808 \log_{10}(T_s/T) \\ & -1.2816 \times 10^{-7}(10^{11.334(1-T/T_s)} - 1) \\ & +8.1328 \times 10^{-3}(10^{-3.49149(T_s/T-1)} - 1) \\ & + \log_{10} e_{ws} \end{aligned} \qquad (2)$$

where

T_s = steam point temperature (373.16°K)

T = absolute (thermodynamic) temperature (°K)

e_s = saturation vapor pressure over a plane surface of pure ordinary liquid water (mb)

e_{ws} = saturation pressure of pure ordinary liquid water at steam point pressure (mb).

Two saturation vapor pressures were calculated for each grid point using (a) the analysis-model field, giving ESAIR, and (b) the prognostic-model field, giving ESX. Then relative humidity parameters were calculated as follows:

$$RHR = \frac{EAIR}{ESAIR} \cdot 100$$

and

$$RHX = \frac{EX}{ESX} \cdot 100.$$

3. <u>Beta Visibility Parameter</u>

The computation of this parameter starts with the production of an extinction coefficient, β, which is a function of windspeed and relative humidity.

$$\beta = F(VVWW) \cdot F(RHR \text{ or } RHX)$$

where VVWW = surface windspeed (m/sec) and

RHR or RHX = relative humidity,

and

$$F(x) = A_1 + x(A_2 + x(A_3 + x(A_4 + x(A_5 + A_6 x)))).$$

If the relative humidity input has a value greater than 99.5 then it is set equal to 99.5.

The coefficients are as follows:

For VVWW < 7 m/sec

	VVWW	RHR or RHX
A_1	0.8065629	-0.4072407×10^{1}
A_2	0.4852030×10^{-1}	0.3865717
A_3	0.5359734×10^{-2}	$-0.1405736 \times 10^{-1}$
A_4	0.0	0.2496362×10^{-3}
A_5	0.0	-0.216801×10^{-5}
A_6	0.0	0.7388672×10^{-8}

For VVWW $\geq$ 7 m/sec

	VVWW	RHR or RHX
A_1	-0.8504248×10^{1}	-0.6135706×10^{1}
A_2	0.3782149×10^{1}	0.583962
A_3	-0.6052896	-0.214833×10^{-1}
A_4	0.4835776×10^{-1}	0.3777016×10^{-3}
A_5	$-0.1915719 \times 10^{-2}$	-0.328404×10^{-5}
A_6	0.3078907×10^{-4}	0.1120986×10^{-7}

Next, a new extinction coefficient is computed as,

$\beta_{TOT} = \beta + S$ where S is given as follows

S	Present Weather Code
0.0	<50
0.35	50-59
0.2	60,61,80
0.6	62,63,81
1.19	64,65,82

The scheme does not apply if weather codes other than those listed above are observed. The weather codes are defined in the Federal Meteorological Handbook No. 2 [U.S. Departments of Commerce, Defense, and Transportation, 1969].

Next, beta visibility is computed by

$$BVISR = \frac{3.91}{\beta_{TOT}} \quad , \text{ using RHR, and}$$

$$BVISX = \frac{3.91}{\beta_{TOT}} \quad , \text{ using RHX.}$$

The modified beta visibility for use with prognostic times is computed without the weather code input by using the formula

$$MBVIS = \frac{3.91}{\beta} \quad ,$$

and here RHX only is used for the relative humidity input.

APPENDIX C

STATISTICS

1. The coefficient of part determination, R^2, may be interpreted as the proportion of the variance of the predictand that is explained by the regression equation. The computation of R^2 follows [Hill, 1979].

Y_i = observed value of the dependent variable for case i.

$\hat{Y}_i$ = regression-specified value for case i

$\overline{Y}$ = mean of the dependent variable

$(Y_i-\hat{Y}_i)$ = residual for case i, also called forecast error

$\sum_i (Y_i-\hat{Y})^2$ = sum of squares about the regression line

$\sum_i (Y_i-\overline{Y})^2$ = sum of squares of deviations about the mean

R = correlation coefficient between Y_i and $\hat{Y}_i$

R^2 = proportion of the variance of Y_i that is "explained" by using $\hat{Y}_i$, or

$$R^2 = \frac{\sum(Y_i-\overline{Y})^2 - \sum(Y_i-\hat{Y}_i)^2}{\sum(Y_i-\overline{Y})^2} .$$

2. The F-to-Enter criterion used to enter variables in the stepwise regression procedure is given as follows [Hill, 1979].

For each independent variable, X_k, that is not in the equation at step (j+1), (j variables have already entered the equation);

F-to-Enter =

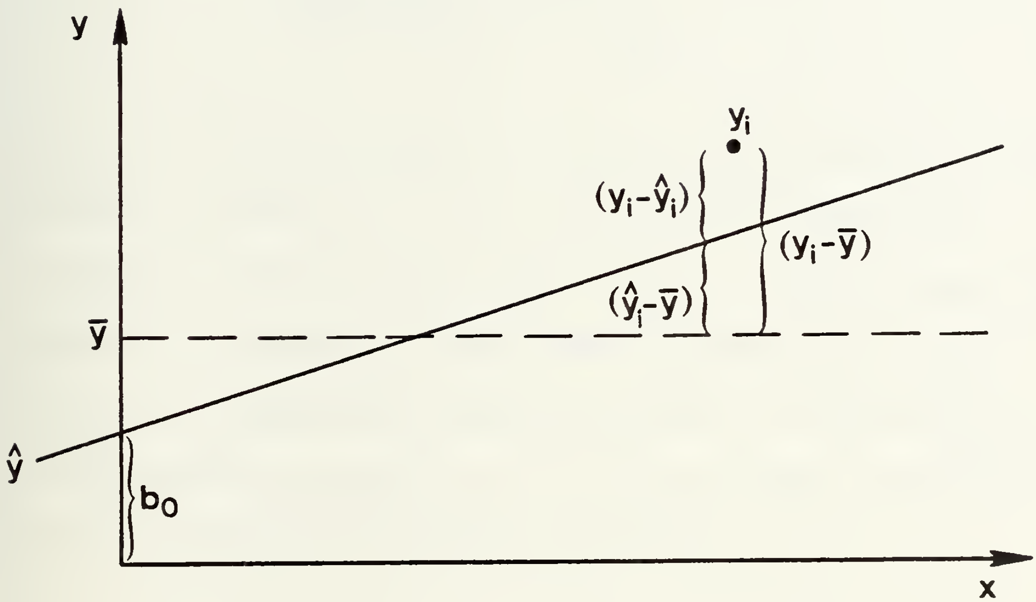
$$\frac{\sum_i (\text{residuals at step } j)^2 - \sum_i (\text{residuals at step } (j+1) \text{ with } X_k \text{ in the equation})^2}{\sum_i (\text{residuals at step } (j+1) \text{ with } X_k \text{ in the equation})^2 / (n-j-2)}$$

n = number of cases

The F-to-Enter statistic is generally a measure of the importance of one variable relative to another.

3. The goal in regression is to find the line, $\hat{Y}$, such that the sum of the squared residuals $[\sum (Y_i-\hat{Y})^2]$ is minimized [Hill, 1979]. For the line to be useful, it is required that the deviations between the observations and the line be smaller than the deviations between the line and the overall mean. Therefore, the quantity

$[\sum (Y_i - \overline{Y})^2] - \sum (Y_i - \hat{Y}_i)^2]$ should be large or one could say a good line has $\sum (Y-\hat{Y})^2$ small compared to $\sum (Y_i - \overline{Y})^2$



The regression line is $\hat{Y} = b_0 + b_1 X$, or generally,

$$\hat{Y}_i = b_0 + \sum_j b_j X_{ji}$$

4. When an independent variable has a low tolerance it should not be included in a regression equation because its value can be expressed fairly well using a linear combination of variables already entered in the equation.

A variable with a low tolerance does not add significantly to the accuracy of a regression equation and may cause numerical and statistical accuracy problems [Hill, 1979]. The tolerance is computed by

$$\text{TOLERANCE} = 1 - R^2_{X_k \cdot \ell}$$

where R is the multiple correlation coefficient of the entering variable, X_k, with the set of independent variables already in the equation, ℓ. If the computed value of tolerance is less than a preselected limit value, a prospective predictor cannot be selected for the regression equation as it is too highly correlated with the predictors already selected.

APPENDIX D

VERIFICATION SCORE FORMULAE

1. The two scores, percent correct and Heidke skill score, use a verification matrix as follows: (A 2x2 matrix is used as an example, but the technique may be applied to any size matrix.)

estimated

observed			
	A	B	i
	C	D	k
	j	ℓ	

$i = A+B$
$k = C+D$
$j = A+C$
$\ell = B+D$

(a) Percent Correct $= \frac{A+D}{A+B+C+D} \times 100$

$$= \frac{\text{number of correct estimates}}{\text{total number of estimates}}$$

(b) Heidke skill score $= \frac{(A+D) - EXP}{(A+B+C+D) - EXP}$

$$= \frac{\text{number of correct estimates - correct number expected due to chance}}{\text{total number of estimates - correct number expected due to chance}}$$

$$EXP = \frac{(i \cdot j) + (k \cdot \ell)}{A+B+C+D}$$

2. Bias Calculation

Bias in estimating a given category =

$$\frac{\text{number of estimates of a given category}}{\text{number of observations of same category}}$$

such as $\frac{j}{i}$ or $\frac{\ell}{k}$.

APPENDIX E

SELECTED VERIFICATION MATRICES

The following verification matrices show the number of observations in relation to the number of regression estimates for each visibility category. The top number in each block is derived from dependent data and the bottom number from independent data. Row and column totals are given in the margins.

1. Verification Matrix for 5P00:

Observed Category		Regression estimated category					
		I	II	III	IV	V	
	I	2	2	174	273	70	521
		8	2	225	293	80	608
	II	4	5	133	231	74	447
		2	1	99	165	60	327
	III	3	2	110	323	150	588
		1	2	105	239	197	544
	IV	1	0	58	299	340	698
		0	1	48	234	408	691
	V	0	0	54	455	1316	1825
		1	0	39	448	2009	2557
		10	9	529	1581	1950	TOTALS
		12	6	516	1379	2819	

2. Verification Matrix for 5P24:

Observed Category	Regression estimated category I	II	III	IV	V	
I	21 13	21 4	111 129	331 337	57 97	541 580
II	13 6	10 5	91 58	269 174	81 68	464 311
III	14 3	4 6	64 54	305 231	201 226	588 520
IV	2 0	4 4	34 33	260 198	398 436	698 671
V	3 3	0 3	31 44	410 350	1360 2088	1804 2488
	53 25	39 22	331 318	1575 1290	2097 2915	TOTALS

3. Verification Matrix for 5P48:

Observed Category		Regression estimated category: I	II	III	IV	V	
	I	2	2	70	365	127	566
		0	0	40	336	171	547
	II	1	2	34	286	143	466
		0	0	24	147	131	302
	III	0	1	33	269	295	598
		0	2	14	193	295	504
	IV	0	0	18	206	461	685
		0	0	9	175	468	652
	V	0	0	17	298	1472	1787
		0	0	14	276	2126	2416
		3	5	172	1424	2498	TOTALS
		0	2	101	1127	3191	

4. Verification Matrix; Probabilistic vs. Categorical

This verification matrix shows results from dependent data for the probabilistic scheme of Aldinger [1979] vs. the 5CAT categorical scheme of this study. The upper values in each block are for the probabilistic scheme, the lower values are for the categorical scheme.

		Regression estimated category					
		I	II	III	IV	V	
Observed Category	I	106	275	139	113	81	714
		7	3	106	504	94	714
	II	76	275	264	198	93	906
		5	2	100	644	155	906
	III	83	284	483	461	141	1452
		2	2	90	902	456	1452
	IV	77	232	380	976	246	1911
		1	1	60	820	1029	1911
	V	117	327	333	2240	1120	4137
		0	1	53	1110	2973	4137
		459	1393	1599	2988	1681	TOTALS
		15	9	409	3980	4707	

LIST OF REFERENCES

Aldinger, W. T., 1979: Experiments on Estimating Open Ocean Visibilities Using Model Output Statistics. M.S. Thesis, Dept. of Meteorology, Naval Postgraduate School, Monterey, CA., 81 pp.

Daniel, C. and Wood, F. S., 1971: Fitting Equations to Data. John Wiley & Sons, 342 pp.

Glahn, H. R., and Lowry, D. A., 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteor., 11, 1203-1211.

Guttman, N. B., 1978: Study of Worldwide Occurrence of Fog, Thunderstorms, Supercooled Low Clouds and Freezing Temperatures. NAVAIR 50-1C-60-CH1, National Climatic Center, EDS, NOAA, Asheville, N.C., 148 pp.

Haltiner, G. J., 1971: Numerical Weather Prediction. J. Wiley & Sons, 317 pp.

Hill, M. A., 1979: Annotated Computer Output for Regression Analysis. BMDP Technical Report, University of California Press, Los Angeles, CA., 76 pp.

List, R. J., 1963: Smithsonian Meteorological Tables. Smithsonian Institution, Washington, D.C., 527 pp.

National Weather Service, 1973: Combined Aviation/Public Weather Forecast Verification. Operations Manual, Chapter C-73, 14 pp.

Nelson, T. S., 1972: Numerical-Statistical Prediction of Visibility at Sea. M.S. Thesis, Dept. of Meteorology, Naval Postgraduate School, Monterey, CA., 36 pp.

Noonkester, V. R., 1980: Offshore Aerosol Spectra and Humidity Relations Near Southern California. Bull. Am. Meteor. Soc., 60, 1272., and Proceedings of AMS Second Conference on Coastal Meteorology, Los Angeles, CA., 113-120.

Ouzts, S. O. and Renard, R. J., 1979: Refinement of a Statistical Diagnostic Model of Marine Fog Using FNWC Model Output Parameters, NPS Report NPS63-79-002, Dept. of Meteorology, Naval Postgraduate School, Monterey, CA., 87 pp.

Quinn, P. F., 1978: Further Development of a Statistical Diagnostic Model of Marine Fog Using FNWC Model Output Parameters. M.S. Thesis, Dept. of Meteorology, Naval Postgraduate School, Monterey, CA., 73 pp.

Schramm, W. G., 1966: Analysis and Prediction of Visibility at Sea. M.S. Thesis, Dept. of Meteorology, Naval Postgraduate School, Monterey, California, 54 pp.

United States Departments of Commerce, Defense, and Transportation, 1969: Federal Meteorological Handbook No. 2, Synoptic Code. NAVAIR 50-ID-2, U.S. Govt. Printing Office, Washington, D.C., 123 pp.

United States Naval Weather Service Command, 1975: United States Naval Weather Service Numerical Environmental Products Manual. NAVAIR 50-1G-522, Dept. of the Navy, Washington, D.C., 339 pp.

University of California at Los Angeles, 1979: BMDP Biomedical Computer Programs, W. J. Dixon and M. B. Brown, editors, University of California Press, 880 pp.

Van Orman, B. L. and Renard, R. J., 1977: Statistical Diagnostic Modeling of Marine Fog Using Model Output Parameters. NPS Report NPS-63Rd77061, Dept. of Meteorology, Naval Postgraduate School, Monterey, CA., 94 pp.

Weaving, W. S., 1977: Screening Regression. Selected Topics in Statistical Meteorology, AWS-TR-77-273. Miller, R. G., editor, Air Weather Service, U.S. Air Force, 143 pp.

Weinreb, M. P., 1971: Comparison of Approximations for Saturation Vapor Pressure of Liquid Water. National Environmental Satellite Service, NOAA, U.S. Dept. of Commerce, Silver Spring, MD., 4 pp.

Wells, W. C., Gal, G. and Munn, M. W., 1977: Aerosol Distributions in Maritime Air and Predicted Scattering Coefficients in the Infrared. J. of Appl. Optics, 16, 654-659.

Wheeler, S. E., and Leipper, D. F., 1974: Marine Fog Impact on Naval Operations. NPS Report 58Wh74091, Dept. of Oceanography, Naval Postgraduate School, Monterey, CA., 118 pp.

BIBLIOGRAPHY

Afifi, A. A. and Azen, S. P., 1972: Statistical Analysis: A Computer Oriented Approach. Academic Press, 366 pp.

Air Weather Service, 1977: Guide for Applied Climatology. AWSP 105-2, U.S. Air Force, 150 pp.

Mosteller, F. and Tukey, J. W., 1977: Data Analysis and Regression. Addison-Wesley Publishing Co., 588 pp.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Air Weather Service Technical Library Scott AFB, Illinois 62225	2
4.	Dr. Robert J. Renard, Code 63Rd Department of Meteorology Naval Postgraduate School Monterey, California 93940	10
5.	Captain Philip G. Yavorsky 114 Trout Avenue Colorado Springs, Colorado 80906	4
6.	Lieutenant W. Tyson Aldinger 47-736-5 Hui Kelu St. Kaneohe, Hi 96744	1
7.	Commanding Officer Fleet Numerical Oceanography Center Monterey, California 93940	1
8.	Commanding Officer Naval Environmental Prediction Research Facility Monterey, California 93940	1
9.	Dr. Donald P. Gaver, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
10.	Dr. Andrzej Goroch Naval Environmental Prediction Research Facility Monterey, Califoria 93940	1

11. Meteorology Reference Center, Code 63 1
Department of Meteorology
Naval Postgraduate School
Monterey, California 93940

12. Captain Wm. J. Kaveney 1
AFIT/CIRF
Wright/Patterson AFB, Ohio 45433

13. Mr. Leo Clarke 1
Fleet Numerical Oceanography Center
Monterey, California 93940

14. Naval Air Systems Command 1
Attn: Cdr K. Van Sickle (AIR-370)
Washington, D.C. 20360

15. Chairman, Department of Meteorology, Code 63 1
Naval Postgraduate School
Monterey, California 93940

16. Chairman, Department of Oceanography, Code 68 1
Naval Postgraduate School
Monterey, California 93940

17. Mr. A. Donnelly 1
Atmospheric Environment Service, AFFF
4905 Dufferin
Downsview, ONT, CANADA

18. Director, Naval Oceanography and Meteorology 1
National Space Technology Laboratories
Bay St. Louis, Mississippi 39520

19. Dr. H. R. Glahn 1
Director, Technique Development Laboratory
National Weather Service
National Oceanic and Atmospheric Administration
Bldg. 8060, 13th Street
Silver Spring, Maryland 20910

Thesis
Y327
c.1

188659

Yavorsky
Experiments concerning categorical forecasts of open-ocean visibility using model output statistics.